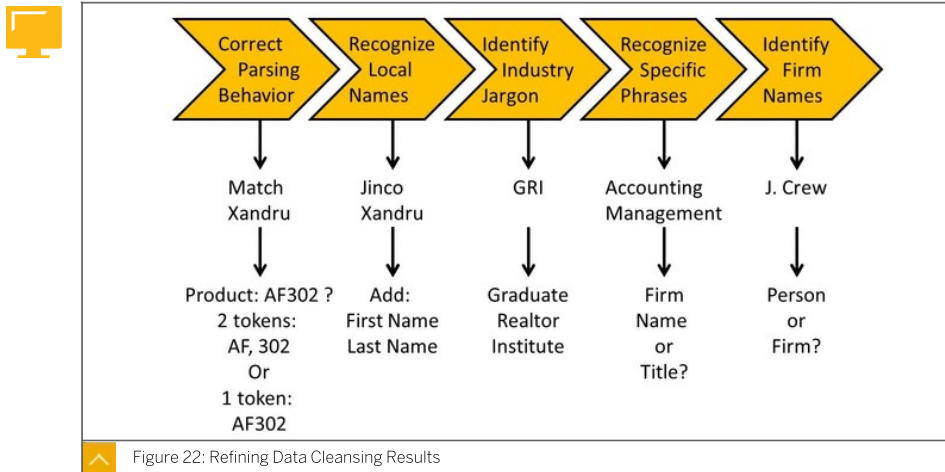




With the introduction of the Cleansing Package Builder in Information Steward, you no longer need to specify individual dictionary, rules, and reference data files. The information formerly contained in those files is now included in the cleansing package. The Cleansing Package option group and Cleansing Package option has been added to the Data Cleanse transform. You can modify the parsing dictionary used by the Data Cleanse transform to improve parsing results for your data.

Improving parsing results:

1. Correct specific parsing behavior using options
2. Recognize local names by adding additional entries
3. Identify industry-specific jargon like special titles
4. Recognize specific phrases
5. Identify firm names containing personal names: for example, is *Johnson* a personal name or part of the firm name *Johnson & Johnson*?

Correct specific parsing behavior

You can customize the parsing dictionary to correct specific parsing behavior that you have seen in your output.

Recognize local names

The name data in Data Cleanse's default parsing dictionary, PERSON_FIRM_EN, is based on an analysis of USA residents. As such, the parsing dictionary is broadly useful across the United States. However, you can tailor the dictionary to better suit your data by adding ethnic or regional names. If Data Cleanse does not recognize a specific name, for example, *Jinco Xandru*, you can add *Jinco* to the dictionary as a first name and *Xandru* as a last name.

Identify industry-specific jargon

The default parsing dictionary is useful across many industries. You can tailor the dictionary to better suit your own industry by adding special titles, prenames or postname, standardized abbreviations, or other jargon words. For example, if you process data for the real estate industry, you might add industry-specific postnames such as Certified Residential Specialist (CRS), Accredited Buyer Representative (ABR), or Graduate Realtor Institute (GRI).